ESSENCE AND COUNTERFACTUALS: COORDINATION AND EXTENSION

Alfredo Watkins

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Philosophy department in the College of Arts and Sciences.

Chapel Hill
2017

Approved by:
Marc Lange
Gillian Russell
Thomas Hofweber
Essence statements are sentences stating what it is to be a certain thing or type of thing; these sorts of statements are in fact quite common. They also interact with counterfactual conditionals in interesting ways. This paper will argue that essence statements sometimes license inferences to counterpossibles, and vice versa. Counterpossibles are counterfactual conditionals where the antecedent posits some metaphysical impossibility. These sorts of counterfactuals are all vacuously true according to the standard Lewis-Stalnaker semantics, even though it seems that some are false (and that some are true, but non-trivially so). If the standard semantics is to account for this, it will have to be modified. As a result, I will propose an alternative understanding of essence and an account of counterfactuals that extends the Lewis-Stalnaker account, and suggest that this combined account might make better sense of some of the phenomena.
For Shannon
# TABLE OF CONTENTS

List of Figures .................................................................................................................. vi  
Section 0: Introduction ..................................................................................................... 1  
Section I: Essence and the Modal Account ...................................................................... 3  
Section II: The Lewis-Stalnaker Account of Counterfactuals ........................................ 8  
Section III: The Coordination of Essence and Counterfactuals .................................... 10  
Section IV: Extending the Framework: The Problem of Counterpossibles .................. 18  
Section V: Adding Impossible Worlds ........................................................................... 25  
Section VI: Real Essence, Similarity, and (Some) Counterpossibles ............................ 27  
  VI.1: Grades of Modality: The Kment Picture .............................................................. 27  
  VI.2: Strict and Narrow Logical Possibility .................................................................. 30  
  VI.3: Real Essence ........................................................................................................ 31  
  VI.4: The Extended Lowe-Kment Account and Essence-Counterpossible Reasoning ... 33  
Section VII: Conclusion ................................................................................................ 38  
Appendix 1: Quantified Anaphoric Counterfactuals ..................................................... 40  
Appendix 2: Would Below and Not-Might Above Inference Patterns ............................. 41  
Bibliography ..................................................................................................................... 42
LIST OF FIGURES

Figure 1 – The Spheres of Modality ................................................................. 29
Figure 2 - The Extended Lowe-Kment Picture .................................................. 34
0. Introduction

Essence statements are sentences stating what it is to be a certain thing or type of thing. These sorts of statements are in fact quite common. They also interact with counterfactual conditionals in interesting ways. In particular, I will argue counterfactuals and essence statements are coordinated according to at least two inference patterns, which I call Would Below and Not-Might Above.

Moreover, I will argue that some instances of these inference patterns allow us to infer counterpossibles from essence statements, and vice versa. Counterpossibles are counterfactual conditionals where the antecedent posits some metaphysical impossibility. According to the standard Lewis-Stalnaker semantics, these counterfactuals are all trivially true. However, it seems that some counterpossibles are false (and that some are true, but non-trivially so). If the standard semantics is to account for this, it will have to be modified. Hence, I will propose an alternative account of counterfactuals that modifies and extends the Lewis-Stalnaker account.

Moreover, in those cases where essence statements allow us to infer counterpossibles, arguably the standard modal account of essence is not able to make sense of this relation of inference. However, I will try to show that when combined with a non-modal account of essence that does not take all of a thing's necessary properties to be essential to it, the validity of these inferences make sense.

In the first part of this paper (Sections I-II) I will explain what essence statements are, at the same time collecting and examining some of the ways that essence locutions arise in different linguistic contexts. The goal of this section is to show that essence language is quite common, and arises in many different contexts. I will also explain the standard “intensional” accounts of essence
and counterfactuals – i.e., the modal conception of essence and the Lewis-Stalnaker semantics for counterfactuals.

After this (in Section III) I will look at some of the ways that essence statements and counterfactuals plausibly interact, in particular arguing for the validity of Would Below and Not-Might Above. I will discuss to what degree the intensional accounts of essence and counterfactuals can make sense of this coordinative relationship. Next, (Sections IV – VI) I will raise the issue of counterpossible conditionals, and show how a modified Lewis-Stalnaker semantics plus a nonmodal account of essence can make sense of at least some of these counterpossibles. Finally, I will briefly conclude with some thoughts on the broader linguistic and philosophical implications of our discussion.
I. Essence

Let’s begin by explaining what we mean by “essence” in the first place. Intuitively, the essence of a thing x or of a kind of thing F is *what it is to be x* or F. An essence statement is a statement of a thing’s essence. For example, a statement of the essence of water might go: “For something to be water is for it to be H2O.”

Note that essence statements can be partial or full. The last example seems to be a full essence statement, in that it appears to give a complete account of what it is for something to count as water. Not all essence statements are like this though; for instance, “It is part of what water is that it be composed of hydrogen.” Here we state a condition that is only part of what it is for something to count as water. But this condition is certainly not a complete account of what it is for something to count as water (lots of non-watery things are composed of hydrogen).

There are several idioms of English that can be used to express essence statements. The following is a (partial, non-exhaustive) list of essence locutions (with “partial” essence locutions on the left and “full” essence locutions on the right):

- “Part of what it is to be F is X” / “What it is to be F is X”
- “Part of the nature of F is X” / “The nature of F is X”
- “Part of what makes something F is X” / “What makes something F is X”
- “For something to count as F it must be X” / “For something to count as F is X”
- “Part of being F *just is* X” / “Being F *just is* X”
- “It is (part) of the (very) essence of F to/that X” / “It is the (very) essence of F to/that X”
• “It is (part) of the (very) definition of F to/that X” / “It is the (very) definition of F to/that X”

• “It is essential to F to/that X” / “It is the essence of F to/that X”

Obviously not every use of these phrases is an expression of an essence claim; but they often are used to express essence claims.

Note that, despite the metaphysical connotations of the word “essence,” these sorts of locutions are not limited to abstract philosophical or metaphysical discussion. Essence claims are made in a variety of linguistic contexts. Let’s consider some examples.¹ One natural starting point is to look at natural science:

a. Physical Science and Chemistry

• (1) Water just is H2O.² (Chemical compounds)

• (2) Having atomic number 79 is what it is to be a gold atom. (Chemical elements)

• (3) Beta decay is when a beta ray and an antineutrino are emitted from an atomic nucleus. (Physical processes)³

b. Biological Science

• (4) “The definition of life is controversial. The current definition is that organisms maintain homeostasis, are composed of cells, undergo metabolism, can grow, adapt to their environment, respond to stimuli, and reproduce.”⁴ (Biological properties)

• (5) Cats are essentially a type of animal. (Biological kinds)

• (6) Meiosis is a process where a single cell divides twice to produce four cells containing half the original amount of genetic information. (Biological processes)

¹ Note that I do not necessarily endorse the truth of each statement below. The point is to get across what essence talk means and how it can be used to make claims.

² This is the famous example due to Kripke in Naming and Necessity (1981) pp. 128 - 129.


⁴ This is an actual quote, taken from the Wikipedia page on Life.
c. Mathematics

- (7) It is part of the definition of a triangle that it has three sides. (Mathematical objects)

Essence locutions are not limited to scientific subjects or contexts, however. They can also occur in more ordinary contexts about more mundane subjects:

d. Societal and Economic Systems

- (8) It is part of the definition of communism that the means of production are controlled by the proletariat. (Economic systems)
- (9) It is part of the very nature of capitalism that there is a free market.

e. Subcultures, Movements, Genres, Styles, and Theories

- (10) DIY (doing it yourself) is of the very essence of punk. (Subcultures)
- (11) That view is no part of our movement! (Social movements)
- (12) Butter is one of the essential ingredients of French cooking. (Styles)
- (13) It is part of the phlogiston theory that combustion releases a massive substance. (Theories)
- (14) A: It is part of the Buddhist worldview that we are each reincarnated. B: Well, I don’t really believe that. A: Well then you don’t really believe Buddhism! B: Hmm – I guess not.

f. Persons and Things

- (15) It is not essential to anyone’s humanity that they have a certain skin color.
- (16) It is essential to me that I was born of the exact parents I was.
- (17) Something could have three legs and still count as a chair.
• **(18)** A: What are you nuts?! That’s not a fridge! B: But look, it can cool things too! [Points to chilled soda in his ice chest.] A: Yeah, but that doesn’t make it a refrigerator!

**g. Philosophy**

• **(19)** To be is to be the value of a bound variable.

This list is not exhaustive; there are many other examples. But the point of this list is to show that essence claims are not limited to either philosophical subjects or philosophical contexts. Given that we have some grasp of essence talk, how should we parse this out more precisely? For the purposes of simplification I will limit myself to characterizing the notion of *partial* essence. Moreover, for now I will limit myself to the two cases of an object and its essential features, and of a kind and its essential features. However, in most cases it should be relatively clear how to extend this reading to other categories.

Now, one natural suggestion\(^5\) is to think of the essential features of a thing as those features it could not exist without, i.e., those features without which it could not be what it is. More precisely:

• **(E1):** Feature F is essential to object a iff it is not possible for a to exist without feature F.

• **(E2):** Feature F is essential to kind K iff it is not possible for something to be a K without having feature F.

Or, in slightly more formal terms:

• **(E1*):** Feature F is essential to object a iff \(\neg \diamond (a \text{ exists } \& \neg Fa) \iff (a \text{ exists } \rightarrow Fa)\]

• **(E2*):** Feature F is essential to kind K iff \(\neg \exists x(Kx \& \neg Fx) \iff \forall x(Kx \rightarrow Fx)\]

This is what I will be calling the modal or intensional account of essence.\(^6\) We will see reasons for thinking this definition is inadequate later on, but for now it is easy to see how it is attractive as a

---

\(^5\) E.g., see example (17) above.
parsing of ordinary essence claims. Aside from its technical elegance, it seems natural to paraphrase essence claims in the way the modal account describes: To say that cats are essentially animals is to say that something could not be a cat and not be an animal. To say that it is not essential to chairs that they have four legs is to say that something could have fewer or more legs and still be a chair. And so on.

\[\text{6 This is the view expressed in Kripke and Putnam’s work, and later developed substantially in Plantinga (1974). See Fine (1994a) for discussion.}\]
II. The Lewis-Stalnaker Account of Counterfactuals

Now that we have some understanding of the notion of essence under our belts, we can turn to how essence statements interact with counterfactual statements. Before jumping straight into that, though, let’s briefly review the standard Lewis-Stalnaker semantics for counterfactuals. Recall that a counterfactual conditional comes in either two basic forms: A would-counterfactual is of the form “If A had been the case, then B would have been the case.” This can be symbolized as “A□→B.” A might-counterfactual is of the form “If A had been the case, then B might have been the case.” This can be symbolized as “A◊→B.”

Often (though not always) the antecedent A will be some sentence that is false in the actual world. Moreover, I will follow Lewis in taking counterfactual conditionals to include both of the subjunctive forms “If A had been the case, then B would have been the case” and “If A were the case, then B would be the case.”

On the standard Lewis-Stalnaker semantics for would-counterfactuals, we evaluate them for truth relative to a possible world as follows:

---

7 In fact, counterfactuals arise in English in a number of forms; see below.

8 This symbolization is taken from Lewis Counterfactuals (1973).

9 Hence, these have sometimes been called “contrary-to-fact” conditionals, e.g., Chisholm (1946). This is inadequate in a number of respects; in particular, since the antecedent does not have to be “contrary-to-fact,” and also because not all conditionals with contrary-to-fact antecedents are counterfactuals (e.g., certain indicative conditionals). Nevertheless, this is a helpful way to think about them.

10 As opposed to only counting the first of these as “counterfactuals.” See Lewis Counterfactuals pp. 1 - 4.
(LS): A counterfactual conditional “A□→B” is true at a world w iff either (i) there is no possible world where A is true, or (ii) there is some possible A&B-world\(^{11}\) that is closer to w\(^{12}\) than any A&¬B-world.

For the sake of simplification, I will take for granted the Limit Assumption about the relation of closeness mentioned in (LS□→):\(^{13}\)

(LA): For any world w and sentence A there is a unique set of A-worlds closest to w. (If there are no possible A-worlds, then this set is the empty set.)

The Limit Assumption says that worlds cannot get infinitesimally close to w without end. Using (LA) we can define a selection function f(A,i) that takes a world i and antecedent A to the set of A-worlds closest to i f(A,i). Again, this function is well-defined because the Limit Assumption is true. In that case, we can reformulate the standard Lewis-Stalnaker semantics as follows:

(LS+): A counterfactual conditional “A□→B” is true at a world w iff either (i) there is no possible world where A is true, or (ii) A&B is true at all of the closest A-worlds, i.e., all of the worlds in f(A,w).

I will also follow Lewis in assuming that might-counterfactuals can be defined in terms of would-counterfactuals as follows\(^{14}\):

A◊→B iff ¬(A□→¬B)

Hence, assuming the Limit Assumption holds, might-counterfactuals have the truth conditions:

A counterfactual conditional “A◊→B” is true at a world w iff (i) there is some possible world where A is true and (ii) A & B is true at some of the closest A-worlds, i.e., some of the worlds in f(A,w).

---

\(^{11}\) An A&B-world is a world where A&B is true. So “There is some A&B-world” means “There is a world u such that A&B is true at u.”

\(^{12}\) I.e., more relevantly similar. See Lewis Counterfactuals p 67 for a discussion of closeness as comparative similarity.

\(^{13}\) Lewis discusses the Limit Assumption in Counterfactuals at pp 19 - 21. Probably the Limit Assumption should not be taken for granted; however, using it makes things technically easier, and its falsity does not much affect what I will have to say.

\(^{14}\) See Lewis Counterfactuals p. 2 and pp. 21 – 24.
III. The Coordination of Essence and Counterfactuals

With a basic understanding of essence and counterfactual statements on the table, let’s look at some of the ways these statements plausibly interact. First, let me state my hypothesis:

- **(CO-K):** Essence and counterfactual statements are coordinated in the following ways:
  
  (i) If feature F is essential to kind K,\(^\text{15}\) then the counterfactual “If something were K then it would be F” is true.

  (ii) If the counterfactual “If something were not F it might still be K” is true, then feature F is not essential to kind K.

- **(CO-O):** Essence and counterfactual statements are coordinated in the following ways:

  (iii) If feature F is essential to object a, then the counterfactual “If a had existed then it would have been F” is true.

  (iv) If the counterfactual “If a had not been F it might have still existed” is true, then feature F is not essential to object a.

I will call (i) and (iii) **Would Below\(^\text{16}\)** and (ii) and (iv) **Not-Might Above.\(^\text{17}\)**

Let us symbolize “Feature F is essential to kind K” by “Ess(F,K).” Similarly, “Feature F is essential to object a” is symbolized by “Ess(F,a).”

It is rather more difficult to parse “If something were K then it would be F” into standard logic. In the first place, the reference of “it” in this case is anaphoric and exhibits (frustratingly) similar behavior to donkey anaphora. Moreover, this subjunctive seems to automatically take us to non-actual worlds and tell us about the Ks in those worlds. Hence, although it might have seemed

---

\(^{15}\) By this I mean that feature F is partly essential to kind K.

\(^{16}\) Since the inference is from an essence statement “above” to a would statement “below”.

\(^{17}\) Since the inference is from a might-counterfactual with a negated antecedent “above,” to an essence statement “below”.
adequate to translate this as “If there had existed Ks then all Ks would have been Fs,” this cannot be adequate, since this counterfactual might be true if there are actually Ks and all of them are Fs,\(^{18}\) even though had there been different Ks they might not have been Fs.\(^{19}\)

Another suggestion is perhaps to translate this as “For all x, if x had been a K then x would have been an F,” where the quantifier “for all x” quantifies into the counterfactual.\(^{20}\) One issue with this is that it automatically commits one to such counterfactuals as “If my copy of *Counterfactuals* had been a dog, it would have had four legs” and “If my house were a butterfly, it would have been delicate.” Perhaps these are fine, but it also isn’t entirely obvious that these are entailed by “If something were a dog, it would have four legs,” and “If something were a butterfly it would be delicate,” etc. More importantly though, this parsing only tells us about the counterfactual features\(^{21}\) of things in the actual world; it does not entail, however, that had there been different dogs, they too would have had four legs.

It is actually not even totally clear how to parse “Had there been different dogs, they would have had four legs.” One might ask “different from what”? Hence, I propose that we treat the antecedent of this counterfactual as positing the existence of dogs different than some world under consideration. We can define “There are different Fs (than at w)” as follows:

- **Diff(x,w)Fx** is true at world u iff there are some different Fs at u than at w., i.e., some Fs that do not exist at w.

In that case, “If there had been different Ks, then all Ks would have been Fs” comes out to have the truth conditions:

---

\(^{18}\) Assuming that strong centering is right, i.e., that the actual world is the closest world to itself.

\(^{19}\) Thanks to Marc Lange (p.c.) for noticing this point.

\(^{20}\) Which is itself an open formula of the form “If x had been K, x would have been F.” See *Counterfactuals* pp. 36 – 43 for a very elegant application of Lewis’s counterpart theory to explain how these open formulae are evaluated relative to a variable assignment.

\(^{21}\) Or as Lewis calls them, “potentialities.” See *Counterfactuals* sec. 1.9
• “[\text{Diff}(x,w)Kx \Box \rightarrow \forall x(Kx \rightarrow Fx)]” is true at world i iff at the world u nearest to i where there are different Ks than w, all the Ks at u are Fs.

With that said, when we ask what would have happened had there been different dogs, we usually mean “What would have happened had there been different dogs than there actually are?”\textsuperscript{22} For the purposes my paper then, I will parse “If something were K it would be F” officially as “If there had been different Ks than there actually are, then all Ks would have been Fs,”\textsuperscript{23} i.e.:

\* (Q) \[ \text{Diff}(x,\@)Kx \Box \rightarrow \forall x(Kx \rightarrow Fx) \]

However, I will often abbreviate this as \((x \text{ is } K \rightarrow x \text{ is } F)\) for convenience.\textsuperscript{24} Then (i) and (ii) are:

\* (i) \(\text{Ess}(F,K) \vdash [(x \text{ is } K) \Box \rightarrow (x \text{ is } F)]\). (Would Below)

\* (ii) \[ \neg(x \text{ is } F) \Diamond \rightarrow (x \text{ is } K) ] \vdash \neg\text{Ess}(F,K). \text{(Not-Might Above)}

Analogous symbolizations for the case of objects are easy to formulate.

This official parsing is still not entirely adequate as an explanation of the meaning of “If something were K then it would be F,” since it only tells us what would happen in the closest worlds where there are different Ks. Nevertheless, it is closer in meaning to the original phrase, and it will still allow us to derive interesting counterpossibles later.

Now, note that Would Below and Not-Might Above are not equivalent to each other. For example, although Not-Might Above allows us to derive: \(\text{Ess}(F,a) \vdash [\neg(a \text{ is } F) \Box \rightarrow \neg(a \text{ exists})]\),

\textsuperscript{22} Though we don’t always \textit{have} to mean this; one can consider a possible world where there are unicorns, and then ask what things would have been like had there been different unicorns than \textit{these} ones.

\textsuperscript{23} Or perhaps “If there had been different Ks than there actually are, then all Ks would have been Fs; and all Ks are Fs.”

\textsuperscript{24} The form \( (x \text{ is } K) \Box \rightarrow (x \text{ is } F) \)” is more amenable to a future treatment that gives a correct syntactic and semantic parsing of “If something were K then it would be F”; it is also much more natural to work with. For now, it can be shown that whenever this sentence is true the corresponding official parsing is true. See Appendix 1.
this, of course, does not entail **Would Below**\(^{25}\); recall that on the standard Lewis-Stalnaker semantics \((\neg X \square \rightarrow \neg Y)\) is not equivalent to \((Y \square \rightarrow X)\).\(^{26}\)

For ease of reference, I summarize all of the forms of **Would Below** and **Not-Might Above**, including both object and kind versions, as well as contraposed forms, in **Appendix 2**.

Now, on the face of it, assuming we have some grasp of what an essential property is, **Would Below** and **Not-Might Above** seem plausible: After all, if \(F\) is an *essential* feature of \(Ks\), then it seems if something were a \(K\) it would have to be an \(F\). Similarly, if it is true that had there been \(Ks\) they might not have been \(F\), then \(F\) cannot be *essential* to being \(K\). The object versions of these inference rules seem plausible for similar reasons. I take this to be some evidence that \((C\O-K)\) and \((C\O-O)\) are right.

Nevertheless, to bolster our case we should consider some examples and see whether \((C\O-K)\) and \((C\O-O)\) accurately predict cases of felicitous and infelicitous reasoning. While there are many examples one can imagine, it is better to look at just a few in detail, setting up the context of the conversation a bit.

**Would Below**

*Example 1: The essence of gold*\(^{27}\)

Consider the following discourse:

- A and B are making various elements in the lab and studying their atomic numbers.\(^{28}\)

  A: Look, we’ve made some platinum -- atomic number 78!

  B: Interesting, I don’t like these grayish metals though. What if we had made some shiny

---

\(^{25}\)Again, **Would Below** says: IF \(\text{Ess}(F,a)\) THEN \([a \text{ exists } \square \rightarrow a \text{ is } F]\).

\(^{26}\)The fact that counterfactuals cannot be contraposed on the Lewis-Stalnaker semantics gives us more possibilities for essence-counterfactual interactions than there would be if they *could* be contraposed!

\(^{27}\)See Boris Kment’s *Modality and Explanatory Reasoning* (2014) at p. 160 for a similar example.

\(^{28}\)They are very lucky scientists; their equipment is quite advanced and reliable.
yellow gold? What would its atomic number have been?

A: It would have had atomic number 79.

B: Huh. How come?

A: Well, that’s just what gold is.

B: Ah, okay. I wish we had made some gold then; I love yellow, and it just so happens 79 is also my favorite number!

This seems to be a perfectly good response on A’s part; A is assuming that we can infer from “having atomic number 79 is (part of) what it is to be gold” to “if something were gold it would have atomic number 79.” This is an instance of Would Below:

- (i) \( \text{Ess}(F,K) \vdash [(x \text{ is } K) \square \rightarrow (x \text{ is } F)] \)

And in general, this seems a perfectly legitimate way of reasoning: If it is part of a kind’s essence that it is F, then if there were anything of that kind it would have to be F.

Note also that if A had told B that atomic number 79 is part of what gold is, it would have been strange for B to have replied: “Okay, got it; but if we had made gold, some of it might also have had atomic number 59, right?” If B had said this, it would have been appropriate for A to insist “No – having atomic number 79 is part of what gold is!” The infelicity on B’s part (and the felicity of A’s response) is predicted by the contrapositive version of Would Below:

- (i* \( \vdash [(x \text{ is } K) \lozenge \rightarrow \neg(x \text{ is } F)] \) \vdash \neg \text{Ess}(F,K) \)

---

29 Actually, the scientist is inferring to something strictly stronger than my official parsing of “If something were gold it would have atomic number 79,” i.e., to the original meaning of the phrase. I take it that this stronger claim nevertheless entails my official parsing.

30 Note that B also seems to be reasoning according to Would Below; he seems to grasp the essence claim, and to infer that if they had made some gold it would have been of his favorite (atomic) number.
Example 2: The nature of democracy

- Suppose A and B are arguing about democracy.

  A: It would be so much better if we lived in a democracy; I hate having to do what other people say. I don’t even know why we have a king; just because he pulled some sword out of a rock shouldn’t make him boss.

  B: Hmm, why would a democracy be so much better? What do you take a democracy to be?

  A: Well, democracy by definition is where everything is voted on by the people.

  B: But that can’t be right; even if we had lived in a democracy, the people might not have voted on everything. Some democracies might allow people to elect representatives, who sometimes vote for them.

  A: Okay, fine. But a democracy would still be better than this; even if we lived in that sort of democracy, we’d at least get to choose who represents us.

B reasons that even if something were a democracy it might have been a representative one; so it can’t be part of the very definition of a democracy that the people vote on everything. This is a perfectly reasonable line of thought; it is an instance of the contrapositive version of Would Below:

- \((i^*)\) \([(x \text{ is } K) \Rightarrow \neg(x \text{ is } F)] \vdash \neg\text{Ess}(F,K)\)

Note, by the way, that when B says “Democracy is X by definition,” he need not be interpreted as stipulating what he means. Although sometimes “by definition” is used this way, it need not be; it is sometimes used to make a claim about what is essential to something, as for example when biologists debate about “the definition of life.” In a discourse like the one we are imagining, the person playing B’s role might well concede that his “definition” is wrong if he is shown that he has gotten the nature of the thing wrong via a counterexample (counterfactual or otherwise). If his definition were a mere stipulation however, it could not be mistaken.

---

31 See example (4) above, on page 5.
**Not-Might Above**

*Example 3: It has to be a hexagon!*

Consider the following discourse:

- A has hired B to make a hexagonal flower arrangement in his garden.
  
  A: So I want the flowers to form a hexagon around the bench.
  
  B: Alright, sure. So would it be alright if I put one row here, one row here, one row here, …
  
  [B traces out five rows in a pentagonal form.]
  
  A: No; if you were to do that it wouldn’t be a hexagon!
  
  B: Huh? Why?
  
  A: A hexagon *by definition* has six sides!
  
  B: Oh, right, sorry; I forgot what a hexagon was!

Here, both A and B correctly reason that, since it is essential to hexagons to have six sides, then if B had not made the arrangement with six sides it wouldn’t have been a hexagon. (A reasons this way at first; B accepts A’s explanation of the counterfactual once he remembers what a hexagon is.) This is an instance of **contrapositive Not-Might Above**:\(^{32}\)

- (ii*) \(\text{Ess}(F,K) \vdash [-\neg(x \text{ is } F) \square \rightarrow \neg(x \text{ is } K)]\)

*Example 4: Height as non-essential*

Consider the following discourse:

- “It’s true that I’m five feet tall, but that can’t be *essential* to who I am. After all, even if I had found some magic beans and turned into a giant, I might *still* have been the same person. In general it doesn’t really matter how tall a person is; it isn’t part of who they are.”

---

\(^{32}\) As a matter of fact, B *also* seems to reason according to **Not-Might Above**: once she realizes her mistake, she agrees based on what a hexagon *is* that if something did not have six sides, it would not be a hexagon.
Again, this seems to be a perfectly legitimate argument: Intuitively, if I might have been the same person even if I were not F, then it can’t be part of my essence that I am F. If it were really part of my very nature to be F then I would have to be F. This is an instance of **Not-Might Above**:

- (iv) \[\neg (a \text{ is } F) \Leftrightarrow (a \text{ exists}) \vdash \neg \text{Ess}(F,a)\]
IV. Extending the Framework: The Problem of Counterpossibles

Intuitively, Would Below and Not-Might Above seem to be perfectly valid modes of reasoning, just on the face of it. And in fact this is predicted by the framework of the standard Lewis-Stalnaker semantics for counterfactuals, combined with the modal account of essence. I will call the combination of these two views the intensional account of counterfactuals and essence.

We can see why this account successfully predicts the validity of Would Below and Not-Might Above. Recall the modal definition of essence:

- (E1*): Feature F is essential to object a iff ¬◊(a exists & ¬Fa) [iff □(a exists → Fa)]
- (E2*): Feature F is essential to kind K iff ¬◊∃x(Kx & ¬Fx) [iff □∀x(Kx → Fx)]

Using the symbols we introduced earlier, we can revise this as follows:

- (E1**): Ess(F,a) iff ¬◊(a exists & ¬Fa) [iff □(a exists → Fa)]
- (E2**): Ess(F,K) iff ¬◊∃x(Kx & ¬Fx) [iff □∀x(Kx → Fx)]

Given the intensional account of counterfactuals and essence, Would Below follows automatically. I will only give the proof for the case of kinds; the case of objects is left as an exercise for the reader.

1. Assume Ess(F,K)

---

33 Again, the intensional account of counterfactuals and essence is the combination of the Lewis-Stalnaker semantics for counterfactuals plus the modal conception of essence.

34 Again, see Appendix 1 for the various versions of Would Below.
2. Then: \( \Box \forall x(Kx \rightarrow Fx) \) [By (E2**)]

3. So, in all possible worlds \( w: \forall x(Kx \rightarrow Fx) \) [This is from the definition of \( \Box \).]

4. Suppose that \( \text{Diff}(x,@)Kx \) is true in no possible world; then \( [\text{Diff}(x,@)Kx \rightarrow \forall x(Kx \rightarrow Fx)] \) is true. [By (LS+)]

5. Suppose that in some possible worlds \( \text{Diff}(x,@)Kx \) is true; then by 3 \( \forall x(Kx \rightarrow Fx) \) is true in the closest such worlds. So \( [\text{Diff}(x,@)Kx \rightarrow \forall x(Kx \rightarrow Fx)] \) is true. [By (LS+)]

6. So either way \( [\text{Diff}(x,@)Kx \rightarrow \forall x(Kx \rightarrow Fx)] \) is true; i.e. \( (x \text{ is } K) \rightarrow (x \text{ is } F) \) is true.\(^{35}\)

Hence, the validity of \textbf{Would Below:} \( \text{Ess}(F,K) \mid \rightarrow [(x \text{ is } K) \rightarrow (x \text{ is } F)] \). Similarly, \textbf{Not-Might Above} follows straightforwardly from the intensional account. Proof:

1. Assume \( \neg[(x \text{ is } F) \rightarrow (x \text{ is } K)] \), i.e., \( \neg[\text{Diff}(x,@) \neg Fx \rightarrow \forall x(\neg F \rightarrow \neg Kx)] \).

2. Then there is some possible world where \( \text{Diff}(x,@) \neg Fx \) is true, and \( \forall x(\neg F \rightarrow \neg Kx) \) is false at some of the closest (possible) \( A \)-worlds. [By (LS+)]

3. So there is a possible world where \( \text{Diff}(x,@) \neg Fx \) is true and \( \forall x(\neg F \rightarrow \neg Kx) \) is false, i.e. where \( \exists x(\neg F \text{ and } Kx) \) is true.

4. Hence, \( \Diamond \exists x(Kx & \neg Fx) \)

5. But \( \text{Ess}(F,K) \rightarrow \neg \Diamond \exists x(Kx & \neg Fx) \) [By E2**]

6. So \( \neg \text{Ess}(F,K) \).

Hence, the validity of \textbf{Not-Might Above:} \( [\neg(x \text{ is } F) \rightarrow (x \text{ is } K)] \mid \rightarrow \neg \text{Ess}(F,K) \)

It is a testimony to the elegance of the intensional account that it predicts the coordination of essence and counterfactual statements in exactly the way (CO-K) and (CO-O) say; this also lends some support to (CO-K) and (CO-O) themselves, insofar as we have independent reason to accept the modal account of essence and the Lewis-Stalnaker account of counterfactuals. Indeed, it is not at

---

\(^{35}\) Recall our conventions of abbreviation; see Appendix 1.
all clear that the Lewis-Stalnaker account of counterfactuals and the modal account of essence were developed specifically with the coordination of essence and counterfactuals in mind.

Unfortunately, in spite of the elegance, simplicity, and intuitive plausibility of the intensional account, there are serious issues insofar as there are recalcitrant data it cannot account for. In particular, its inability to handle certain hyperintensional phenomena will be our primary concern.

In the first place, the modal account of essence entails that any property a thing (or kind of thing) has in all possible worlds is an essential property of that thing (or kind of thing). On the face of it, this seems implausible: Although Socrates necessarily has the property of being such that $2 + 2 = 4$, this does not seem to be part of what it is to be Socrates. Moreover, the modal account immediately entails that any property a necessary being has is an essential property of that thing. But this also seems implausible; even if God and mathematical objects are both necessary beings, it does not seem to be part of what it is to be the number 2 that God is thinking about it.

On the other hand, the standard Lewis-Stalnaker semantics notoriously has the issue that it makes all counterpossibles come out trivially true. Counterpossibles are counterfactual conditionals where the antecedent posits some metaphysical impossibility. Now, on the standard Lewis-Stalnaker semantics (simplifying a bit) a counterfactual $A \Box \rightarrow B$ is true iff either (i) there is no possible $A$-world, or (ii) $B$ is true at all the closest possible $A$-worlds. Hence if $A$ is impossible, then by clause (i) $A \Box \rightarrow B$ is automatically trivially true. But this seems implausible. Intuitively, at least some counterpossibles are false; for instance “If intuitionistic logic were correct, LEM would fail” seems true (though not trivially true) and “If intuitionistic logic were correct, then LNC would fail”

---

36 So it often goes with our theories about the world.

37 See Klima (ms) p. 5. See also Kment (2014) and Fine (1994a).

38 Indeed, it seems we can infer this from the fact that if there had been no God there might still have been mathematical objects; but never mind that for now!

39 Or, in the Stalnaker (1968) semantics, the antecedent $A$ is only true at the impossible world $\lambda$. 

20
seems false; but if intuitionism is false, then it is necessarily false, and so both of these counterfactuals would have to be trivially true according to the Lewis-Stalnaker view. And there are many other cases that have been discussed in the literature.\footnote{Some of these arguments are particularly strong; see especially Jenny (2016). See also Brogaard and Salerno (2013), Yagisawa (1988), Nolan (1997), Berto (2013), and Hofweber (2016) at p. 340 fn. 35.}

In what follows, then, I will assume that there is some evidence against the modal conception of essence, and some evidence that there are counterpossibles that are false and others that are non-trivially true. Hence, aside from the fact that it seems dogmatic to assume counterpossibles must all be trivially true, there is some positive reason for us to be open to the evidence that follows.

However, rather than focus on the issues of essence and counterpossibles individually, I will focus specifically on what appear to be perfectly good counterpossible instances of \textit{Would Below} and \textit{Not-Might Above}. This strategy seems potentially fruitful in a number of ways.

First, insofar as we have independent reason to think \textit{Would Below} and \textit{Not-Might Above} are valid in general, we have some reason to think they hold in the hyperintensional case. In that case, if there are any false counterpossibles of a certain form, this gives us reason to think the modal account of essence fails.\footnote{Proof: Suppose \(\neg\Box\exists x (Kx \& \neg Fx)\), but there is some false counterpossible C of the form (x is K) \(\square \rightarrow (x \text{ is } F)\). Then it follows (x is K) \(\diamond \rightarrow \neg(x \text{ is } F)\), and it follows by \textit{Would Below} that \(\neg \text{Ess}(F,K)\), even though \(\neg\Box\exists x (Kx \& \neg Fx)\), contra the intensional account of essence.} It also shows that counterpossibles can play a “non-trivial” role in our modal and explanatory reasoning, adding further to the evidence that some counterpossibles are non-trivially true and others false.\footnote{See footnote 30 above.} Second, if we can give an hyperintensional account of essence and counterfactuals that validates and explains these inferences, then that is some evidence in favor of that account. Third, if we assume \textit{Would Below} and \textit{Not-Might Above} are valid in the hyperintensional case then this places some restrictions on our account of essence and counterfactuals, and gives us guidance as to how we might develop these accounts together.
In what follows then, I will consider a few examples of counterpossible reasoning that appear to instantiate Would Below and Not-Might Above. These involve counterfactuals whose antecedents are metaphysically impossible. Here I will help myself to examples drawn from contexts that are admittedly philosophical and abstract. To that extent, my conclusions hereafter will be relatively modest and tentative, given the special nature of the examples I’m relying on; nevertheless, the following seem to me to be among the most straightforward and illuminating cases.

**Would Below**

*Example 1: Atheistic reasoning about God*

Consider the following discourse:

- A: To be honest, I want to believe in a God, but one of the biggest problems for me is the existence of evil.

  B: I don’t see the problem; if there were a God he would allow the sorts of evils we see here.

  A: That’s impossible. After all, it is part of the *very definition* of a God that he be *perfectly good*; so if he really existed he wouldn’t allow any evil at all.

  B: But maybe a perfectly good God would have to allow certain evils for the sake of some greater good?

  A: That seems contradictory to me. After all, God is also by nature all powerful, so he wouldn’t need to use evil to bring about a greater good; he could just bring it about directly.

This sort of discourse is perfectly natural. Moreover, the reasoning on A’s part seems perfectly legitimate. On the one hand, A doesn’t believe in God. In fact, she believes that God’s existence is “contradictory.” Nevertheless, she infers from the nature of God that, *were* there a God he would not allow evil, and *were* there a God he would be all powerful. These are both instances of **Would Below**:

---

43 Whether or not we agree with her arguments.
• \( \text{Ess}(F,a) \vdash [(a \text{ exists}) \Box \rightarrow (a \text{ is } F)] \)

Note moreover that A takes herself to be disagreeing with B when B says “If there were a God he would allow the sorts of evils we see here,” i.e., she takes this counterfactual to be false, even though she believes the antecedent to be metaphysically impossible. And standing back, this seems intuitively plausible anyway: Even if there is no God, it doesn’t seem true both that “if there were a God he would allow evil” and “if there were a God he would allow no evil.”

**Not-Might Above**

*Example 2: Euthyphro Dilemma*

Another theological example: Consider the following discourse:

• “I do believe that God always wants us to do the right thing. But even if there were no God, some actions might still have been wrong. Hence, it cannot be part of the very definition of wrongness that it is commanded by God; for even if there were no God to command them, some things might still have been wrong.”

This line of reasoning seems plausible enough. Even if there had been no God, some things might still have been wrong, so it can’t be part of the definition of right and wrong action that God commands it. Note that this line of reasoning instantiates **Not-Might Above**:

• \( \neg(x \text{ is } F) \Diamond \rightarrow (x \text{ is } K) \vdash \neg\text{Ess}(F,K) \)

This seems reasonable in spite of the fact that “morally right” and “commanded by God” are taken by our reasoned to be necessarily coextensive, pace the modal account of essence.

---

44 Witness the famous quote from prominent Dutch jurist Hugo Grotius: “What we have been saying [about morality] would have a degree of validity even if we should concede [etiamsi daremus] that which cannot be conceded without the utmost wickedness, that there is no God, or that the affairs of men are of no concern to him.” See Neff (2012).

45 Though we might disagree with the premise; one way to do so would be to show there is some essential connection between rightness and wrongness and the commands of a lawgiver. This line of reasoning would be in accordance with **Not-Might Above**.
Example 2: Nominalism and Socrates

Consider the following line of reasoning, which could reasonably be uttered by someone who believes mathematical objects are necessarily existing entities. For what it is worth, I am not the first to use this example; it is discussed extensively (and endorsed) by Alastair Wilson (who uses a somewhat different approach to counterfactuals than the one taken here)46:

- “Even if nominalism about sets were true, Socrates might still have existed. Hence, it can’t be part of what it is to be Socrates that he belong to his singleton set.”

The mode of reasoning here involves positing an impossible scenario where there are no set-theoretical objects; but it seems plausible that, even in that case, one could coherently suppose Socrates to still exist. But in that case, it does not seem it can be part of Socrates’ very identity that certain sets exist.47 (This is further evidenced by the fact that both Platonists and nominalists believe, seemingly coherently, in Socrates.) This is an example of Not-Might Above:

- \[ \neg(a \text{ is } F) \Diamond \rightarrow (a \text{ exists}) \vdash \neg\text{Ess}(F,a) \]


47 Note, by the way, that this line of reasoning seemed perfectly plausible when the antecedent was metaphysically possible. For instance: “Even if my skin color had been different, I might have still existed / Hence, it’s not part of who I am that I have a certain skin color.”
V. Adding Impossible Worlds

We have seen what appear to be perfectly good hyperintensional instances of Would Below and Not-Might Above; more precisely, we have seen cases where essence facts seem to allow us to infer to counterpossibles that are true, but non-trivially so, and cases where certain counterpossibles allow us to infer that some property is non-essential to a thing (or kind of thing) even where that property is necessarily had by that thing (or kind of thing). On the intensional account, this should be impossible: All necessary properties of a thing are essential to it, and all (would-)counterpossibles are trivially true (and all might-counterpossibles trivially false).

To make sense of this then, we will have to modify the intensional account. The following is one proposal for how that account might be modified. I will not spend a large amount of space justifying every step; hereafter, my account should be taken as tentative, and plausible to the extent that it makes sense of the phenomena we have looked at and is likely to make sense of further phenomena.

To begin, we must modify the Lewis-Stalnaker semantics so that it does not make all counterpossibles come out trivially true. The most popular way to do this is by considering worlds that are not metaphysically possible and where the antecedent holds. Hence, in our extended Lewis-Stalnaker semantics, we will begin by simply removing the restriction to possible worlds:

- (E-LS): A counterfactual conditional “A□→B” is true at a world w iff there is some A&B-world that is closer to w than is any A&¬B-world.

Moreover, if we take the Limit Assumption for granted, we can simplify this as follows:
• **(E-LS+):** A counterfactual conditional “A□→B” is true at a world w iff A&B is true at all of the closest A-worlds, i.e., all of the worlds in f(A,w).

As before, a might-counterfactual A ◊→ B is defined as ¬(A□→¬B); hence:

- A might-counterfactual “A ◊→ B” is true at a world w iff A&B is true at some of the closest A-worlds, i.e., some of the worlds in f(A,w).

Now, to help get the well-definedness of the selection function f(A,i) we assume:

- **Hyperintensional Space:** For any set of sentences S there is a world where exactly the sentences of S hold.\(^{48}\)

Although some authors have taken the impossible worlds to satisfy various logical principles,\(^ {49}\) our formulation allows us to use the impossible worlds for a wider range of phenomena\(^ {50}\) and seems to be the most natural development anyway, though not much turns on this.

---

\(^{48}\) See Rasmussen (2016) at p. 151.

\(^{49}\) E.g., Priest (2005) and Bjerring (2013) impose various restrictions on how the logical connectives behave at impossible worlds. This seems somewhat arbitrary. See also Berto (2013) for discussion.

\(^{50}\) E.g., it may allow us to assess more counterlogicals than we otherwise would be able to.
VI. Real Essence, Similarity, and Countermetaphysicals

1. Grades of Modality: The Kment Picture

Given our space of impossible worlds then, in order to evaluate our counterpossibles we need to order some of the impossible worlds according to similarity. But given the inherent lack of structure in the hyperintensional space, it may initially seem a daunting task to try to impose some similarity relation on it.

For this reason, my project here will be limited: To see whether some counterpossibles can be put on a solid footing by extending the standard Lewis-Stalnaker semantics. I will not try to give an account that might explain every sort of counterpossible one might come across. I will limit myself to trying to explain the sorts of counterpossibles we’ve examined, along with the corresponding hyperintensional instances of Would Below and Not-Might Above.

To start with, I will need the notion of different grades of modality. I will build on an account suggested by the work of David Lewis and elaborated by Boris Kment. I will not give an extended defense of this account of the various grades of necessity and possibility here. Rather than try to reinvent the wheel in just a few pages, I will simply lay out what the view is, and see whether there is a natural extension of this view that perhaps allows us make sense of some of the counterpossibles we’ve examined, along with the instances of Would Below and Not-Might Above. To that extent then, my argument can be taken as conditional and modest: If the Lewis-

---

51 Though I am optimistic about our prospects for explaining many different sorts of counterpossibles.

52 See Lewis Counterfactuals sec. 2.5 pp. 52-56.

53 See Kment Modality and Explanatory Reasoning Ch. 2 -3, 7.
Kment picture of the grades of modality is correct, then perhaps my extended account can make sense of some of the counterpossibles we are interested in.

Let us begin with the notion of a similarity sphere:

- A sphere $S$ around a world $i$ is a set of worlds such that every world $u$ in $S$ is closer (more similar) to $i$ than any world $v$ outside of that sphere.

Now, Lewis suggests that we can understand comparative possibility as corresponding to comparative similarity with the actual world. For example, to say that (A) “I am one inch taller” is more possible than (B) “I am the size of the Empire State Building” is to say that there is some world where I am one inch taller which is more similar to the actual world than any world where I’m the size of the Empire State Building; in other words, there is some sphere around the actual world that contains an A-world, but no B-world.\(^{54}\)

Boris Kment builds on this suggestion and uses the notion of comparative possibility to define the intuitive notion of grades of possibility and necessity.\(^{55}\) Kment hypothesizes that the various grades of modality, such as physical possibility and metaphysical necessity, correspond to certain degrees of possibility and necessity, i.e., certain spheres around the actual world. In particular, Kment takes the set of physically possible worlds to be the sphere around actuality that matches actuality with respect to all of the physical laws, and the set of metaphysically possible worlds to be the sphere around actuality that matches actuality with respect to all of the metaphysical laws.\(^{56}\) The set of physical (metaphysical) necessities can then be defined as those

---

\(^{54}\) See Lewis *Counterfactuals* sec 2.5.

\(^{55}\) See Kment *ibid* p. 32; see ch. 2 sec. 4 more generally.

\(^{56}\) Cf. Kment *ibid.* See also Kment ch. 7.
truths that hold throughout all of the physically (metaphysically) possible worlds. I will call this view the **Kment Picture of Modality**. The following picture illustrates this conception:

![Fig. 1: The Spheres of Modality](Note: Figures not drawn to scale.)

Initially, the Kment Picture may seem rather uninformative, since understanding what is physically or metaphysically possible apparently presupposes some grasp of what the physical and metaphysical laws are. However, to demand more than this would be to expect too much from an abstract semantic account of the grades of modality: For it is not the *job* of semantics to tell us what the actual laws are; the job of *semantics* is to show how our various modal notions are systematically related. Intuitively, however, plausible candidates for physical laws include the laws of our best physics; plausible candidates for metaphysical laws include truths about the nature of causality, truths of mereology, truths about the essences of things, mathematical truths, and so on.

Note that on the Kment Picture it immediately follows:

- **(S-P)** Any physically possible world is closer to actuality than any world that is not.
- **(S-M)** Any metaphysically possible world is closer to actuality than any world that is not.
Though it is beyond the scope of this paper to test this extensively, it does seem to explain certain intuitive counterfactual judgments, like “If I had dropped this pen, it would have fallen”\(^{57}\) and “If a tree had come into being, it would have had a cause.”\(^{58}\)

2. Strict and Narrow Logical Possibility

In order to make sense of the sorts of counterpossibles we’ve discussed, I propose that we extend this picture even further to include grades of modality beyond the metaphysically possible. The plan for the rest of the paper is this: Building on a suggestion of E.J. Lowe’s, I will acknowledge the existence of two additional grades of modality that include the metaphysically possible: strict logical possibility and narrow logical possibility. I will explain this in intuitive terms without defending it in any detail. I will then suggest that we further acknowledge a sphere of worlds that is included within these spheres, and that itself includes the metaphysically possible worlds, viz. the set of worlds where all the essential truths hold. Hence, I will briefly explain how we should think of the essential truths. We will call this set of worlds the essence compatible worlds. My hypothesis is that, with an appropriate understanding of the essential truths, we can use this machinery to make sense of some of the counterpossibles and inference patterns that we looked at earlier. I will call the extended structure on hyperintensional space that I develop the Extended Lowe-Kment Picture of Modality. First, Jonathan Lowe:

“In the previous section I was content to characterize logical possibility somewhat vaguely in terms of compliance with the laws of logic, with logical necessity understood correspondingly. In fact, one can distinguish three different grades of logical necessity, as follows. First there is the strictly logically necessary – that which is true in virtue of the laws of logic alone. Secondly there is the narrowly logically necessary – that which is true in virtue of the laws of logic together with definitions of non-logical terms. And thirdly there is the broadly logically necessary … one might very reasonably contend that this last grade of logical

\(^{57}\) Assuming this follows from the actual laws of physics; e.g., laws about gravity.

\(^{58}\) Assuming it is a metaphysical law that whatever begins to exist has a cause.
necessity is in fact coextensive with metaphysical necessity – indeed, that they are just two different names for the same thing.59

Though I will not follow Lowe’s discussion exactly, I believe these distinctions are helpful. We can think of the strictly logically possible worlds as those worlds where the laws of logic hold. Intuitively, this may be broader than the class of metaphysically possible worlds, since there seem to be metaphysical necessities that are not themselves laws of logic or true in virtue of the laws of logic. For instance, even if it is metaphysically necessary that every contingent fact has an explanation,60 it is presumably not true just as a matter of logic that this is so. Similarly, the class of narrowly logically possible worlds seems to extend more widely than the class of metaphysically possible worlds; for although it is metaphysically necessary that water is H2O, it is not true either in virtue of the laws of logic alone, nor is it true by definition.61

I will simply assume this picture is right and that there are grades of modality beyond the metaphysically possible. I will next identify as a sphere included within the narrowly logically possible worlds what I will call the set of essence compatible worlds, i.e., the sphere where all the essential truths hold. The next task then is to briefly lay out what these essential truths are, especially if they are not merely de re metaphysical necessities as the modal account suggested.

3. Real Essence

As we saw earlier, the modal account of essence suffered from a few prominent flaws. Recall what the modal account said:

- \((E1**)\): \(\text{Ess}(F,a) \iff \neg \Diamond (a \text{ exists } \& \neg F(a)) [\iff \Box (a \text{ exists } \rightarrow F(a))]\)


60 This is a version of Leibniz’s famous Principle of Sufficient Reason; it is a candidate for a metaphysical law.

61 If one has doubts about the notion of “truth in virtue of definitions” or “analytic truth,” then most of what I say can be reformulated sans the notion of narrow logical possibility without much consequence.
(E2**): \( \text{Ess}(F, K) \) iff \( \neg \exists x (Kx \& \neg Fx) \) [iff \( \forall x (Kx \rightarrow Fx) \)]

One problem with this definition was that it made any logical consequence of “Fx” part of a thing’s essence; hence, on the modal account, it was part of what it is to be Socrates that \( 2 + 2 = 4 \). Also, on the modal account, any property of a necessary being is essential to that thing; hence, it is an essential property of the number 2 that God is thinking about it.

Given these problems with the modal account, I do not expect that we will be able to fully reduce essence statements into non-essentialist (e.g., modal) terms. We will likely have to work with the intuitive notion of an essential feature of an object (or kind) as one that is part of what it is to be that object (or kind). Nevertheless, I think we can say a little bit more about the structure of essence statements than we already have.

In the first place, I propose that we think of essential truths as being relative to an object or a given class of objects.\(^{62}\) On my view, an essential truth is a truth that holds in virtue of what it is to be a given thing or type of thing. So, for instance, “water is H2O” is an essential truth about water. And “if Socrates exists he is human” might be an essential truth about Socrates. These are both essential truths, because it is part of what it is to be water that water is H2O, and it is part of what it is to be Socrates that he is human. Hereafter, I will think of the essential truths as being prefixed by an operator: Where \( p \) is an essential truth about \( X \)’s, we can say “it is true in virtue of the essence of \( X \) that \( p \)” or “\( [E(X)]p \)”\(^{63}\)

Note that some essential truths give necessary conditions and others give sufficient conditions.\(^{64}\) So for instance, plausibly, it is an essential truth about gold that anything with atomic number 79 is gold; this is a sufficient condition. On the other hand, it is plausibly an essential truth

\(^{62}\) See Kment (2014) ch. 6. See also Fine 1994b.

\(^{63}\) Kment takes a similar approach; see Kment (2014) ch. 6.

\(^{64}\) See Kment (2014) pp. 147 – 159.
about cats that anything that is a cat is an animal; this is a necessary condition. Some essential truths may give both necessary and sufficient conditions.\textsuperscript{65}

With this parsing of the language, I will analyze the locutions we have given earlier in these terms. In particular, I will take essence statements of the form “F is essential to a” and “F is essential to K” as follows:

- \( \text{Ess}(F,a) = \text{df} \ E(a) \) (if a exists, a is F)
- \( \text{Ess}(F,K) = \text{df} \ E(K) \) (if Ks exist, they are all F)

Plausibly, essential truths are themselves metaphysically necessary. Moreover, although we cannot define the essential truths in terms of modal facts, it does seem plausible that if something has a feature essentially, then it has it in all metaphysically possible worlds where it exists:\textsuperscript{66}

- \( \text{(E1***): } \text{Ess}(F,a) \) only if \( \neg \Diamond (a \text{ exists } \& \neg Fa) \) [only if \( \Box (a \text{ exists } \rightarrow Fa) \)]
- \( \text{(E2***): } \text{Ess}(F,K) \) only if \( \neg \Diamond \exists x (Kx \& \neg Fx) \) [only if \( \Box \forall x (Kx \rightarrow Fx) \)]

Because of this, where the antecedents of the relevant counterfactuals are metaphysically possible, \textbf{Would Below} and \textbf{Not-Might Above} are still valid.\textsuperscript{67} Hence, in the case of metaphysically possible antecedents, this theory of essence is able to account for the same predictions as the intensional account; in the case of counterpossibles, we hope to do even better.

\textbf{4. The Extended Lowe-Kment Account and Essence-Counterpossible Reasoning}

With some grasp on what the essential truths are, it is intuitive to see that they may well not include all of the metaphysical laws. For instance, even if it is a metaphysical law that everything that begins to exist has a cause, it does not immediately seem to be true in virtue of what it is to be any

\textsuperscript{65} Perhaps corresponding to the notion of “complete” as opposed to “partial” essence statements.

\textsuperscript{66} See Fine (1994a) p. 8.

\textsuperscript{67} See the proof on pp. 17 - 18 above.
sort of thing or type of thing. Or even if it is a metaphysical law that any two objects constitute a third\(^\text{6}\) it isn’t clear that this law holds in virtue of the nature of any particular object or kind of objects. Hence, I propose that we think of the essential truths as constituting a sphere around actuality that includes, but is not contained in, the set of metaphysically possible worlds:

![Diagram of Lowe-Kment Picture](image)

**Fig. 2: The Extended Lowe-Kment Picture (Note: Figures not drawn to scale.)**

- The sphere of *essence compatible* worlds $= \text{df}$ the set of all *strictly logically possible* worlds where the essential truths hold.

Of course, if this is a sphere, then the following immediately follows:

- \((\text{S-E})\) Any strictly logically possible world where all the essential truths hold (i.e., any *essence compatible world*) is closer to actuality than any world where they do not.

This hypothesis gives us what I have called the **Extended Lowe-Kment Picture of Modality**. We will use this proposal to see whether we can make sense of some of the counterpossible/essence

---

\(^6\) As the laws of mereology hold.
reasoning that we saw earlier. With our technical machinery now in place, let us return to the hypothesis about essence-counterfactual coordination considered earlier:

- **(CO-K):** Essence and counterfactual statements are coordinated in the following ways:
  
  (v) If feature F is essential to kind K, then the counterfactual “If something were K then it would be F” is true.
  
  (vi) If the counterfactual “If something were not F it might still be K” is true, then feature F is not essential to kind K.

- **(CO-O):** Essence and counterfactual statements are coordinated in the following ways:
  
  (vii) If feature F is essential to object a, then the counterfactual “If a had existed then it would have been F” is true.
  
  (viii) If the counterfactual “If a had not been F it might have still existed” is true, then feature F is not essential to object a.

We saw several plausible instances of these patterns of reasoning in Sections III and IV above. Now, however, I would like to qualify this hypothesis. As we will see, we will need to restrict CO-K and CO-O to certain contexts. Once they are restricted to these contexts, however, I will suggest that CO-O and CO-K give us valid inference rules, even when the antecedents of the counterfactuals in question are metaphysically impossible.

But first, some evidence that we should include a restriction: Consider the case of a contradictory object, such as a square circle, or a circle C that is defined to be both round and not round.69 Perhaps there are some essential truths about these objects. For instance:

- (C1) It is part of the essence of C that it is round.
- (C2) It is part of the essence of C that it is not round.

---

69 There are worlds where such objects exist, by our characterization of hyperintensional space on p. 25.
Then, by C1 and **Would Below** it follows that if C had existed it would not be round. By **Would Below**, we can infer it is not part of the essence of C that it is round, which together with (C1) forms a contradiction. Similar problems arise with **Not-Might Above**.

Maybe these examples are not convincing, but rather than take a stand on this issue, I will simply restrict CO-O and CO-K to cases where the relevant antecedent holds in some essence compatible world (which, recall, is a subset of the strictly logically possible worlds). In that case, given the Extended Lowe-Kment Picture, plus our account of essence, we can actually derive the revised version of CO-O and CO-K. Proof (I will only give the object version):

- **(xi) (Would Below):** Suppose F is essential to object a. Then in any world where all the essential truths hold, a is F if a exists. Then in any essence compatible world where a exists, a is F. Now, suppose that the existence of a is essence compatible. Then there is some essence compatible world where a exists; in any such world, a is F. But, by the Extended Lowe-Kment Picture, any such world is closer than any essence incompatible world. Hence, in the closest world where a exists, a is F. Hence, if a had existed, a would have been F.

- **(xii) (Not-Might Above):** Suppose that even if a were not F, a might still have existed. Suppose that a’s not being F is essence compatible, i.e., there is some essence compatible world where it is not the case that a is F. By our assumption, there is some closest world where a fails to be F and a exists. Hence, in that world, it is not true that if a exists then a is F. But by the Extended Lowe-Kment account, this world is among the essence compatible worlds. Hence, it is not an essential truth that if a exists then a is F. Hence, feature F is not part of the essence of a.

Note that these antecedents may well be metaphysically impossible. We can see how this works in a couple of the cases we mentioned earlier:

- **Case 1:** It is part of the essence of God that he is perfectly good. Hence, if there were a God, he’d be perfectly good.

Assuming that it is essence compatible that there be a God, i.e., the existence of God does not violate any essential truth (although it may be metaphysically impossible), then in the closest worlds where there is a God, God has all his essential features (by the Extended Lowe-Kment Picture).
Hence, if it really is part of what it is to be God that he is perfectly good, then if God had existed he’d have been perfectly good.

- **Case 2**: Even if nominalism about sets were true, Socrates might have existed. Hence it is not part of the essence of Socrates that he belongs to his singleton set.

By the premise, in some of the closest worlds where nominalism about sets is true, Socrates exists. But assuming that the non-existence of the sets does not violate any essential truth (even though it is *metaphysically* impossible) by the Extended Lowe-Kment Picture it follows that these worlds are among the essence compatible ones. Hence, there is an essence compatible world where Socrates exists and there are no sets; in particular, there is an essence compatible world where Socrates exists and his singleton set does not exist. Since there is an essence compatible world where “if Socrates exists his singleton set exists” is false, it follows that it is not essential to Socrates that he belong to his singleton set.
VII. Conclusion

One can see how this might apply to other cases too: For instance, it seems false that if I’d been a soul, I’d have been a material being (since it is part of what it is to be a soul that it is immaterial); and it seems true that had there been no abstract objects, the material world still might have been there (since the material world doesn’t depend for it to be what it is, on the cold, distant Platonic horde); or had the A-theory of time held gold would still have had atomic number 79.

On the other hand, I have deliberately limited myself to explaining and proposing the modified forms of CO-O and CO-K (restricted to antecedents that are essence compatible). I certainly have not given a general account of all counterpossibles; for instance, I have nothing to say at this point about counterfactuals like “Had there been a round square, it might not have been round.” There are also counteressentials, such as “If communism were a form of capitalism, it would have had free markets” or given a device C that makes things with atomic number 59 there is “If C had made gold, some gold would have had atomic number 59.” I will not try to say much about these here, though perhaps the fact that these antecedents are at least within the range of the strictly logically possible makes them liable to further explanation.

70 I am assuming, perhaps wrongly, that materialism is necessarily true.

71 Though I hold out hope for our being able to say more about these sorts of cases in the future.

72 Thanks to Marc Lange (p.c.) for this nice example.

73 For what it is worth, these cases seem perfectly non-vacuous to me.
I will close by summarizing a few important parts of our discussion. In the first place, it
should be clear from the collection of cases in Section I that essence locutions are in fact a
ubiquitous part of natural language, and that they are not confined to abstract metaphysical or
philosophical contexts. Moreover, we have examined some plausible inference patterns between
essence statements and counterfactuals that seem to hold in general. We have then looked at a few
plausible instances of these patterns where the relevant counterfactuals have metaphysically
impossible antecedents, but which nonetheless seem non-vacuous. This led to the development of
an hyperintensional account of essence and counterfactual statements that allows us to make sense
of some of these counterfactuals.

Of course, I fully admit that I have made a number of tenuous assumptions along the way,
and have built extensively on work already done by others. To the extent that these assumptions are
questionable, my theory is also open to question, and I acknowledge the tentative nature of my
discussion. Nevertheless, what we have seen is that, with the help of a modified notion of essence, it
is perhaps possible to extend one standard picture of counterfactuals to plausibly make sense of
certain counterpossibles and certain essence-to-counterpossible inference patterns. This may provide
some support for the idea that we should acknowledge as perfectly meaningful certain
hyperintensional pieces of natural language, where the hyperintensionality is not merely
“subjective”74 but instead corresponds to an “objective” feature of the world.75

---

74 As in the case of oblique transitive verbs, control verbs, propositional attitude verbs, that introduce hyperintensional
oblique contexts; e.g., “believes that.”

APPENDIX 1: QUANTIFIED ANAPHORIC CONDITIONALS

- **Diff(x,w)Fx** is true at world u iff there are some different Fs at u than at w., i.e., some Fs at u that do not exist at w.

Let us define > as a two place operator from pairs of predicates to formulas:

- \( > \overset{df}{=} \lambda P\lambda Q[\text{Diff}(x,@)P_x \square \to \forall x(P_x \to Q_x)] \)
- \( (F > G) = \overset{df}{=} \lambda P\lambda Q[\text{Diff}(x,@)P_x \square \to \forall x(P_x \to Q_x)](F)(G) \)
  iff \( [\text{Diff}(x,@)F(x) \square \to \forall x(F_x \to G_x)] \)
- \( (F \overset{m}{>} G) = \overset{df}{=} \neg(F > \neg G) = \neg[\text{Diff}(x,@) \square \to \forall x(F_x \to \neg G_x)] \)
  iff \( \neg[\text{Diff}(x,@) \square \to \neg Ex(F_x \& \neg G_x)] \)

Anaphoric literals:

- **x is F** =df \( \lambda xF_x; \)
- **\neg (x is F)** =df \( \lambda x\neg F_x \)

Anaphoric counterfactuals:

- **(x is F) □→ (x is G)** =df \( (x is F) > (x is G) \)
- **(x is F) ◊→ (x is G)** =df \( (x is F) m> (x is G) \)
- **If something had been F then it would have been G** =df \( (x is F) □→ (x is G) \)
- **If something had been F then it might have been G** =df \( (x is F) ◊→ (x is G) \)

Then it can be shown that the official parsing of the relevant counterfactual will come out true if and only if the relevant anaphoric counterfactual is true.
(Would Below)

- (i) \( \text{Ess}(F, K) \models [(x \text{ is } K) \Box \rightarrow (x \text{ is } F)] \)
- (i*) \( [(x \text{ is } K) \Diamond \rightarrow \neg(x \text{ is } F)] \models \neg\text{Ess}(F, K) \)
- (iii) \( \text{Ess}(F, a) \models [(a \text{ exists}) \Box \rightarrow (a \text{ is } F)] \)
- (iii*) \( [(a \text{ exists}) \Diamond \rightarrow \neg(x \text{ is } F)] \models \neg\text{Ess}(F, a) \)

(Not-Might Above)

- (ii) \( [\neg(x \text{ is } F) \Diamond \rightarrow (x \text{ is } K)] \models \neg\text{Ess}(F, K) \)
- (ii*) \( \text{Ess}(F, K) \models [\neg(x \text{ is } F) \Box \rightarrow \neg(x \text{ is } K)] \)
- (iv) \( [\neg(a \text{ is } F) \Diamond \rightarrow (a \text{ exists})] \models \neg\text{Ess}(F, a) \)
- (iv*) \( \text{Ess}(F, a) \models [\neg(a \text{ is } F) \Box \rightarrow \neg(a \text{ exists})] \)

---

76 This is the contraposed form of (i); in general the starred formulas are contrapositives of the unstarred formulas.
BIBLIOGRAPHY


